21.1: CLT-based Inference for Difference in Proportions

Stat 120 | Fall 2025

Prof Amanda Luby

APM Research Lab ran a survey of 800 likely Minnesota voters between Sept 16-18 of this year. They reported 48.4% were in support of the Harris/Walz ticket, and 43.3% were in support of the Trump/Vance ticket.

Here's an excerpt from their methodology report:

> *The margin for error, according to standards customarily used by statisticians, is no more than ± 3.5 percentage points. The margin of error is approximately 7 percentage points for the difference between two data points…. The margin of error is higher for any subgroup, such as gender or age grouping.*

(a) What is the *standard error* for the difference in proportions?

(b) Explain why we should *not* use the CLT formula for the difference in proportions between Harris supporters and Trump supporters in this poll

(c) Instead, we'll look at the difference in suppport for Harris between two different levels of education. (We can assume that the poll included 436 in the "no college degree" group and 360 in the "college degree" group). Show two ways that the standard error could have been computed.

(d) The two ways of computing the standard error should produce similar results. Give two reasons why.

(e) The following R code makes a confidence interval and performs a hypothesis test for the difference in these two proportions.

i. Where did `x = c(192, 194)` come from?

ii. What is α for the hypothesis test?

iii. Provide an in-context interpretation of the confidence interval (it should be clear in your answer which group had more support for Harris)

```r
prop.test(x = c(192, 194),
          n = c(436, 360),
          correct = FALSE, #tells R not to apply a "continuity correction"
          conf.level = .90)
```

```
    2-sample test for equality of proportions without continuity correction

data:  c(192, 194) out of c(436, 360)
X-squared = 7.663, df = 1, p-value = 0.005637
alternative hypothesis: two.sided
90 percent confidence interval:
 -0.15680365 -0.04024018
sample estimates:
   prop 1    prop 2
0.4403670 0.5388889
```

(f) What concept or topic is your group *most* confused about? (Some ideas to get you started: `qnorm` vs `pnorm`, normal vs t distribution, choosing a parameter, standard error formulas, when to use confidence intervals vs hypothesis tests)